Review of “Operational algorithm for ice/water classification on dual-polarized Radarsat-2 images” by Natalia Zakhvatkina and 4 co-authors

General comment: The task of developing robust algorithms for operational sea ice classification is important for obvious reasons. The process of development includes also interesting scientific questions, in most cases related to image processing techniques. Recently, the combination of different polarizations has gained increasing attention because of radar systems capable of acquiring data at two or more polarizations at the same time.
The use of dual-polarization is the main topic of the present manuscript and of topical interest not only with respect to Radarsat-2 but also to ESA’s Sentinel-1 mission. I am in particular impressed about the fact that more than 2700 Radarsat-2 images were classified and compared to operational sea ice maps provided by Met. Norge. This enabled the authors, e. g., to judge the classification accuracies under summer conditions. The paper is clearly structured. However, several passages in the text need to be clarified concerning their content, and the English should be corrected. The largest problem I see is the lack of information about the criteria which were applied to decide which combinations of textural features (out of a group of 9) are optimal for separating sea ice and open water (separately for HH- and HV-polarization), and why?
I recommend accepting this paper after major revisions. I will not provide suggestions for improving the language nor will correct write errors but leave this to the (co-)author(s).

Detailed questions:

(1) page 1, line 24: what is meant by “ice edge dynamic distribution” – is it “(dynamic) ice edge variations”?

(2) page 2, lines 1-2: “…based on surface roughness and other characteristics of the scene”. Surface roughness is an ice property but not a scene characteristic (like texture). What you probably mean is “…based on gray tone variations (caused by variations of the backscattered radar intensity) and other…”

(3) page 2 lines 4-5: should be rephrased…”…and open water can have similar σ^0, depending on wind speed and direction…”

(4) page 2, line 7: “…image texture and OTHERS…” – one more example for “others”?

(5) page 2, lines 27-31: (a) It is not true that grey ice and multi-year ice show similar radar brightness at HH-polarization in any case – usually multi-year and first-year (level) ice can be very well distinguished at C-band HH-polarization. Only during summer the differences between young and old ice diminish. (b) When does new ice appear bright? Usually it appears dark! The situations for a bright appearance that come into my mind are frost flowers or broken and deformed new ice (which due to deformation reveals a very rough surface). (c) At HV-polarization smooth thin ice appears dark, and open water appears dark as well, independent of the wind speed. The examples (a) – (c) have to be corrected and/or to be explained more in detail.

(6) page 2, lines 31-34: Does already the first sentence (“The dual polarization…”) refer to the work by Geldsetzer and Yackel, 2009, or to another study?

(7) The goal of the present study is well described on page 3, lines 20-23. However, in my opinion it would be very useful if the authors emphasize the special motivation of their work, considering the fact that studies on classification with dual-polarization data were already published earlier.

(8) page 4, lines 6-8: This is an important point of your work! You should emphasize here in addition that you have to analyze a larger number of radar images because the radar intensity contrast between open water and ice is varying significantly dependent on ice conditions and wind speed/direction (the latter affecting the radar brightness of open water).

(9) page 4, lines 13-14: Sentence: “The reason is…” I recommend discarding this sentence. Actually, the magnitude of the received cross-polarized signal depends on the structure of the illuminated target. In case of sea ice, e. g., the cross-polarized return increases with increasing macroscopic ice deformation.

(10) page 4, lines 14-16: “This causes…” Does the low intensity at HV-polarization explain its variation in across-track direction? I don’t think so. I guess the reason is technical limitations of the radar sensor electronics when operated close to their noise level. Please check the Radarsat-2 manual. A good description of the HV-variations can also be found in Komarov & Barber, TGRS, Vol. 52, No. 1, pp. 121-136, 2014

(11) page 4, line 22: start “Our AUTOMATED…” Your automated algorithm does not include 6 steps but only 3 as you yourself explain on top of page 5. What you describe following this sentence is the first step of your analysis, namely the (manual) determination of thresholds and suitable textural measures that are later used in the fully automated (unsupervised) classification. Please formally separate the training/test part and the subsequent automatic classification more clearly.

(12) page 4, line 29-30: Do you mean “Training of the automatic classifier (e. g. SVM) using different combinations of texture features together with radar intensity, based on manually classified images”?
(is it really a “training” or rather a determination of thresholds between ice and water?)

(13) page 5, line 5: how large is the sliding window? (You can give a hint to section 3.4 in which you mention the sizes).

(14) section 3.1: I recommend that you provide the equation used for the incidence angle corrections and define the “linear-trend coefficient” that you use later. You should also give a hint that you discuss problems with this correction in section 5.1.

(15) page 5, line 24: “incontinuity” = discontinuity? Since the problematic zones are masked out in the HV-images, one should also see them in the figures that you present later. But masks are not included in the figures.

(16) page 6, sentence lines 2-4: I assume that all co-authors understood this sentence but I do not. “…trained classifier…” = ice analyst from Met. Norge? If the images show, e. g., cracks, ridges and leads, why can’t the “trained classifier” not identify them (they are distinct features)?

(17) page 6, lines 4-6: please separate the single sub-classes more clearly. Is it 1: young ice, first-year ice and multi-year ice, 2: fast ice, 3 broken ice mixed with ice-free water? What is the reason to group the ice types like that? Usually, e. g., young ice and multi-year ice reveal very different radar signatures (at least in intensity, not necessarily in GLCM textures). Younger ice can be mis-classified as open water at lower wind speeds, multi-year ice as open water at higher wind speeds, if one focuses on radar brightness.

(18) page 6 line 14: “the full range” refers to the number of grey tones, which are rescaled to a lower number of bits? How many bits do your original images have?

(19) page 7, lines 5-8: this is another very important part of your work which needs to be described a little more in detail. (a) You vary the computational parameters: window sizes, co-occurrence distances, and quantized grey levels (please mention the increments for the latter two). With these different computational parameters you calculate the different texture features - which means that you have a huge number of possibilities. By which method did you determine the optimal combination of computational parameters (which you mention on page 8 lines 25-30)? Is it

described in your TGRS paper from 2013, to which you refer to in section 4.2? The TGRS paper, however, does not cover all aspects of the processing described here. (b) Which criteria did you apply for deciding which combination of texture features is optimal for classification at HH- and at HV-polarization? (In your TGRS-paper you used all texture features.) (c) How many radar images and related ice charts from Met. Norge did you use for these tests? (d) Were the selected combinations of texture features best for all images, or only in a majority of the investigated cases?

(20) Section 3.6: For ice chart production, also optical images were employed. In how many cases could they not be used because of dense cloud cover? Did you use a "weight" indicating the reliability of the ice charts (assuming that the lack of optical information causes more difficulties for the ice analyst to separate ice and water)?

(21) Page 8, line 15: what do you mean by "…and increases from top right to bottom left"? Do you refer to the extent of the open water area?

(22) Page 8, lines 17-18: Does new and thin ice really always appear brighter than OW at HV-polarization? I doubt this.

(23) Section 4.1: As already noted above (comment 15): How did you handle the computation of texture features in the zones along the borders between the image stripes that you indicated as white lines in 3b?

(24) Section 4.2: Please see comments (19). From my point of view it is up to you whether you provide more details about methodological aspects in section 3.4 or in this section 4.2.

(25) Page 9, lines 3-5, Fig. 4: The usefulness of the different texture features is not clear to me. When combining the intensities at HH and HV, calm OW is very well separated, but OW and ice partly overlap. The latter is also valid for the two graphs to the right, but here even calm OW overlaps wit the ice. Please provide some convincing arguments why the texture features improve the water-ice separation? (See also comment 33 below).

(26) Referring to Fig. 6, 7, 8: why is Fig. 5 not mentioned before them?

(27) Section 4.4: How was the comparison between classified radar image and Met. Norge ice charts carried out? Were the latter digitized to the pixel size of your radar images?

(28) page 9 line 28-30: (a) "…ice charts WERE obtained manually…"; (b) what do you mean by "higher rate of thematic processing" in this context?
I think that also ice charts from Met.Norway could include a higher level of details if the ice analyst delimits small-scale features. But this would require a longer time for ice chart production. What is the processing (computer) time needed when applying your algorithm compared to the semi-automated chart production by Met. Norge? This aspect is important and should be added to the discussion section 5.

(29) Section 5.1: (a) The title refers only to the incidence angle correction, but the noise correction is also dealt with, at least in the last sentence.

(30) page 10, lines 5-6: "Strong dependence of the backscatter on incidence angle in open water surface… is significantly higher than for sea ice". This statement is not true in any case, since the incidence angle sensitivity depends on the surface roughness: the smoother the surface, the larger the sensitivity (in particular close to the angle of specular reflection). Thin level ice can be very smooth and may show a larger sensitivity than a wind-roughened water surface.

(31) (a) page 10, line 8: what do you mean by "overestimated signatures"? (b) Line 9: what do you mean by "ice cover has more reliable backscatter ranges for various ice types"? Since wind speed

can be determined from radar measurements over open water, this means that also open water has clear defined ranges of backscattering intensity for a given wind speed and direction (except for OW patches within the ice cover and at the ice margin where ocean surface wave interactions are more complicated).

(32) Section 5.2: (a) You should make clear that in the first sentence you refer to the nine texture parameters given by equations 2-10. If you in addition use intensity and standard deviation, there are in total 22 parameters considering both HH and HV (and not 20). In the following text you should make clear that you now refer to section 4.2, in which you selected 5 textural parameters + intensity + standard deviation for HH, and 4 textural parameters + intensity for HV, leaving in total 12 parameters. (b) I do not understand the meaning of the two sentences from line 20 to line 23. E. g., what are "poor features"? What is a discriminant function in this context, and why is it needed? If all features are used, does this mean that the classification accuracy is lower in any case? There is of course an optimal number and an optimal selection of textural features giving highest classification accuracy, but with, e. g., only 4 textural parameters you might theoretically still get better accuracies than with less parameters. The last 3 sentences of section 5.2 refer to the reduction of dimensionality but a direct link to the own processing described in section 4.2 is missing (e. g. the criteria for selecting some of the nine given textural features and excluding others).

(33) Section 5.3: With this section I had considerable difficulties. (a) I suggest that you provide a table giving the function of each textural parameter (e. g. measure of local variations), and the interpretation of the respective parameter related to Fig 5.
(b) In this section you introduce alternative denotations for the textural parameters (e. g. energy – angular second moment; homogeneity – inverse difference moment etc). You should make this clear by expressing it like "Energy (also called 'Angular Second Moment')…" (after this, it is fine to go on with e. g. "Homogeneity or Inverse Difference Moment".
(c) page 10 line 30. I found that the "Energy" is the square root of the Angular Second Moment. Is "repeatability" another denotation for energy? In Fig, 5i, the OW area, which appears homogenous in Fig. 3b, reveals a very low energy – is it a noise effect? The bright blue zones in the OW-area in Fig. 5i are due to the stripes in Fig 3b?
(d) page 11 sentence lines 6-7: I do not understand this sentence. The GLCM correlation function is calculated for a co-occurrence distance of 8 pixel? Fig. 5j: I do not understand why the correlation is very high in the marginal ice zone and low in the more closed ice. I would expect it vice versa. The low correlation value over open water is again a noise effect?
(e) page 11, line 13: homogeneity – why is the homogeneity high in the marginal ice zone and low in the inner ice zone and over water (Fig. 5k)? Again I expected this vice versa.
(f) page 11, line 17: "…indicates a random mixture of scattering mechanism". I think here the entropy from the entropy-alpha decomposition described by Cloude et al. (where entropy is indeed related to the character of the scattering mechanism) is mixed with the GLCM entropy, which does not give any information about the scattering characteristics from within one pixel but relations between neighboring pixels.
(g) page 11, lines 26-28 and lines 30-33, regarding Fig. 4: I do not see a clear separation between OW (dark blue) and sea ice (green), and the separation capability of the texture features seems to be worse than the one of the intensities. See also comment 25 above.

(34) Section 5.4, first paragraph: (a) it should be considered that the maps drawn by the ice analysts are a "smoothed version" of the ice cover variations. In principle the ice analysts could also provide more detailed maps. However, I anticipate that an automated algorithm can do this much faster – see also comment 28 above. (b) why didn't you use a better land mask? (c) page 12 lines 17-20: how did you treat the beam boundaries when calculating the texture features (see also comment 23 above)? (d) page 12, lines 32-33: why didn't you exclude cases in which the temporal difference between manual and automatic ice chart was large?